

ORIGINAL CONTRIBUTIONS

Differential Misclassification and the Assessment of Gene-Environment Interactions in Case-Control Studies

Montserrat García-Closas,¹ W. Douglas Thompson,² and James M. Robins^{1, 3}

In case-control studies of interactions between genetic and environmental exposures, differential misclassification of the environmental exposure with respect to disease status can introduce spurious heterogeneity of the stratum-specific odds ratios. In this paper, the authors identify conditions under which differential misclassification does not introduce bias in the interaction parameter when no multiplicative interaction is present, and it biases the interaction parameter toward the null value when a multiplicative interaction is present. The conditions are that (i) conditional on potential confounders, the environmental exposure is independent of the genotype among the controls, and (ii) misclassification of the environmental exposure is nondifferential with respect to the genotype. These conditions can be tested from the misclassified data in the control group, since a test of the independence of the genotype and the misclassified environmental exposure among the controls is a test of the joint hypothesis that conditions (i) and (ii) are both true. Therefore, the authors propose a two-step test for interaction which first tests conditions (i) and (ii) and then goes on to test for interaction, provided the first step hypothesis is not rejected. A summary test procedure to test for gene-environment interactions in the presence of misclassification, based on both a conventional test for interaction and the two-step test, is recommended, and is illustrated with data from a case-control study of the role of diet as a modifier of the association between a metabolic polymorphism and lung cancer. *Am J Epidemiol* 1998;147:426-33.

case-control studies; epidemiologic methods; misclassification

Recent developments in molecular techniques have enabled epidemiologic studies to evaluate the role of genetic markers on disease occurrence, as well as potential interactions between genetic and environmental exposures. The case-control design is frequently used since the outcomes studied are usually rare and the assessment of individual genotypes requires the use of expensive laboratory techniques. One of the major sources of bias in case-control studies is misclassification of subjects with respect to environmental exposures in the past. The classification probabilities are often different for cases and controls due

to their different abilities to recall past exposures or to recent changes in behavior related to disease status, giving rise to differential misclassification with respect to disease (1). Although the assessment of genotypes is not affected by this type of misclassification, the assessment of environmental exposures which could modify the gene-disease association is affected. This differential misclassification can introduce bias on the estimation of the interaction parameter, i.e., the ratio of stratum-specific odds ratios, either toward or away from the null (2). Several investigators have developed methods for correcting for the effects of misclassification of an effect modifier (2-4). However, these methods require accurate estimates of stratum-specific misclassification probabilities, which are often not available.

In this context, it becomes important to identify conditions under which we could predict the effect of exposure misclassification on the stratum-specific odds ratios. In this paper, we show that under conditions often satisfied in studies of gene-environment interactions, differential misclassification of the environmental exposure 1) does not introduce bias in the

Received for publication June 10, 1996, and accepted for publication August 27, 1997.

Abbreviations: CI, confidence interval; *GSTM1*, glutathione-S-transferase; OR, odds ratio.

¹Department of Epidemiology, Harvard School of Public Health, Boston, MA.

²Applied Medical Sciences, School of Applied Science, University of Southern Maine, Portland, ME.

³Department of Biostatistics, Harvard School of Public Health, Boston, MA.

Reprint requests to Dr. Montserrat García-Closas, Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPN/443, 6130 Executive Blvd., Bethesda, MD 20892-7374.

ratio of stratum-specific odds ratios when no multiplicative interaction is present, and 2) it bias the ratio of the stratum-specific odds ratios toward the null value when a multiplicative interaction is present. Moreover, these conditions themselves can be tested using the misclassified data in the control group. This suggests a two-step test in which homogeneity of the stratum-specific odds ratios is tested only if the conditions ensuring the validity of the test in the presence of misclassification are not rejected in the first step. This paper focuses on the effects of misclassification of the environmental exposure and assumes that the genotype is measured without error.

The remainder of the paper proceeds as follows: In the next section, we present the two main results of the paper. We then propose and discuss a two-step test for gene-environment interactions in the presence of misclassification, and give recommendations for testing. These recommendations are illustrated with the example of a case-control study of the modification of the association between metabolic polymorphisms and lung cancer by diet. The last section presents a summary and conclusions.

RESULTS

Consider a case-control study in which the goal is to determine whether the association between a particular genotype and the risk of disease, as measured by the disease-genotype odds ratio, changes according to different categories of an environmental exposure. This is equivalent to assessing whether there are different environmental exposure effects in each genotype category. The term "environmental exposure" is used to denote any nongenetic exposure. We will use the terms "multiplicative interaction," "effect modification," and "heterogeneity of the stratum-specific odds ratios" interchangeably in the remainder of the paper. The terms "interaction parameter" and "ratio of the stratum-specific odds ratios" will also be used interchangeably. In this section, we will assume that there is no bias due to selection or confounding, thereby allowing the use of unadjusted odds ratios. We will further assume that misclassification is only present in data obtained on the environmental exposure, and that the disease and genotype data are correctly classified.

Genotype will be denoted as G with levels $g = 1, 0$, disease will be denoted as D with levels $d = 1, 0$, the true environmental exposure as E with levels $e = 1, 0$, and the misclassified environmental exposure as E' with levels $e' = 1, 0$, where 1 indicates present and 0 absent. Finally, $OR_{DG|E=e}$, $OR_{DE|G=g}$, and $OR_{EG|D=d}$ will represent the conditional odds ratios. Suppose that the following conditions are true:

Condition (i): The true environmental exposure is independent of the genotype among the controls,

$$\text{Prob}(G = g | E = 1, D = 0) =$$

$$\text{Prob}(G = g | E = 0, D = 0), \text{ for } g \in (1, 0).$$

Condition (ii): Misclassification of the environmental exposure is nondifferential with respect to the genotype among cases and controls, i.e., conditional on the true environmental exposure and the disease status, the measured environmental exposure E' with levels $e' = (1, 0)$, is independent of the genotype,

$$\text{Prob}(E' = e' | E = e, G = 1, D = d) =$$

$$\text{Prob}(E' = e' | E = e, G = 0, D = d),$$

$$\text{for } e, e', d \in (1, 0).$$

Then:

Result 1: The estimated interaction parameter will be unbiased in the presence of possible differential misclassification of E , if multiplicative interaction is not present. Consequently, a test for multiplicative interaction will be valid.

Result 2: The estimated interaction parameter will be biased toward the null value in the presence of possible differential misclassification of E , if a multiplicative interaction is present (given that the usual assumptions for nondifferential misclassification biasing the odds ratio toward the null are satisfied (5-7)).

Result 1 can easily be shown to be true if we express the interaction parameter as the ratio of the stratum-specific odds ratios conditioning on disease status rather than conditioning on the environmental exposure, i.e., $OR_{GE|D=1} / OR_{GE|D=0}$. If there is no multiplicative interaction, i.e., $OR_{GE|D=1} / OR_{GE|D=0} = 1.0$, and G and E are independent among the controls (condition (i)), i.e., $OR_{GE|D=0} = 1.0$, then G and E must also be independent among the cases, i.e., $OR_{GE|D=1} = 1.0$ (8). In the presence of misclassification of E that is nondifferential with respect to G (condition (ii)):

$$OR_{GE|D=0} = 1.0 \Rightarrow OR_{GE|D=0} = 1.0$$

$$\text{and } OR_{GE|D=1} = 1.0 \Rightarrow OR_{GE|D=1} = 1.0.$$

Therefore, under the stated conditions $OR_{GE|D=1} / OR_{GE|D=0} = 1.0$, i.e., the estimated interaction parameter will be unbiased in the presence of misclassification of E .

From the above, it follows that the null hypothesis of homogeneity in the absence of misclassification of E , $H_0: OR_{GD|E=1} / OR_{GD|E=0} = 1.0$ implies the null hypothesis of homogeneity in the presence of misclassification of E , $H'_0: OR_{GD|E'=1} / OR_{GD|E'=0} = 1.0$. Therefore, an alpha-level test of H'_0 will be an alpha-

level test of H_0 . Result 1 can be generalized to the case of multilevel exposure since, under conditions (i) and (ii), $\text{Prob}(G | D = d, E' = i) = \text{Prob}(G | D = d) = \text{Prob}(G | D = d, E = i)$.

Another consequence of conditions (i) and (ii) is that, if there is no multiplicative interaction, then the exposure-specific odds ratio for the G - D association can be estimated without bias despite the presence of differential misclassification of E . This follows because, given $\text{OR}_{GD|E=1} = \text{OR}_{GD|E=0}$, they both must equal OR_{GD} since by condition (ii), E and G are independent among the controls (3). Further, we have by result 1, that $\text{OR}_{GD|E'=1} = \text{OR}_{GD|E'=0}$. Therefore, they must also equal OR_{GD} since by conditions (i) and (ii), E' and G are independent among the controls. On the other hand, the estimated genotype-specific odds ratios for the E' - D association, $\text{OR}_{E'D|G=0}$ and $\text{OR}_{E'D|G=1}$, will be biased toward the null value. Since we have assumed that the exposure error is non-differential with respect to G (condition (ii)), the relative bias will be the same for both levels of G , which is the reason why the ratio of stratum-specific odds ratios will still be unbiased.

Result 2 can be shown following a similar rationale. If the true odds ratio for the E - G association among the cases differs from 1.0, then non-differential misclassification of E with respect to G will bias this odds ratio toward the null, provided the usual assumption that sensitivity is $\geq 1 - \text{specificity}$ (7). Thus, the interaction parameter will be the ratio of an attenuated odds ratio to an odds ratio that was and continues to be 1.0, and, therefore, this ratio will also be attenuated toward the null value. Finally, it should be noted that, contrary to result 1, result 2 only applies to the case of a dichotomous exposure and does not generalize to a multilevel exposure.

Misclassification of exposure that is non-differential with respect to disease can, in general, also bias the estimation of the interaction parameter away from the null. Indeed, one can start with a true scenario where there is no interaction but where interaction is introduced by a non-differential misclassification of exposure. However, if, as under our conditions (i) and (ii), one of the odds ratios (i.e., $\text{OR}_{GE|D=0}$) is 1.0, such bias away from the null cannot, as we have proved, occur, even if there is differential misclassification of E with respect to D , as long as misclassification of E is non-differential with respect to G .

IMPLICATIONS FOR DATA ANALYSIS

Result 1 suggests that, in the assessment of gene-environment interactions, we should first consider whether conditions (i) and (ii) are reasonable assumptions in our study and, if we conclude so, perform a

conventional test for homogeneity or a test for homogeneity based only on cases as described by Piegorsch et al. (8). However, we are actually in a better situation, because the conditions for validity can be tested in the control group using the misclassified data. The central insight is that the joint hypotheses that conditions (i) and (ii) are both true (denoted as H_{01}^*) implies independence between the misclassified environmental exposure E' and the genotype G among the controls (denoted as H_{01}). Therefore, any alpha-level test of H_{01} is an alpha-level test of H_{01}^* , i.e., valid tests of H_{01} are valid tests of H_{01}^* . It should be noted, however, that testing H_{01} may not be as informative as one might hope because of low power of the test, since H_{01}^* can be false when H_{01} is true.

Moreover, conditional on H_{01} being true, a test of independence of E' and G among the cases (denoted as H_{02}) is a valid test of hypothesis H_0 of homogeneity of the stratum-specific odds ratios, assuming that conditions (i) and (ii) are true. This is clear if we think of a test for homogeneity as a test of whether the odds ratios measuring the E' and G association for cases and controls are equal, i.e., $\text{OR}_{GE'|D=0} = \text{OR}_{GE'|D=1}$, since conditional on $\text{OR}_{GE'|D=0} = 1.0$ (i.e., independence of E' and G among the controls), a test for homogeneity is a test of $\text{OR}_{GE'|D=1} = 1.0$ (i.e., independence of E' and G among the cases) (8). Because the tests of H_{01} and H_{02} are independent, the alpha-level of the second test does not need to be adjusted based on the outcome of the first test.

In short, when conditions (i) and (ii) are satisfied, we can perform a valid test for homogeneity in the presence of misclassification by conducting two independent tests for the genotype-exposure association, the first a test of the hypothesis H_{01} among the control subjects, and the second a test of the hypothesis H_{02} among the cases. This two-step procedure is depicted in figure 1.

Following this sequence of independent tests, the interpretation of the type I errors in each step would be as follows. If conditions (i) and (ii) are true, then:

Step 1: With probability α_1 we will reject the hypothesis that conditions (i) and (ii) are true and make no decision concerning the homogeneity of the odds ratios. With probability $(1 - \alpha_1)$ we will proceed to step 2.

Step 2: Conditional on proceeding to step 2 and on hypothesis H_0 of homogeneity of the stratum-specific odds ratios being true, with probability $(1 - \alpha_2)$ percent we will correctly accept the hypothesis H_0 of homogeneity. With probability α_2 percent, we will falsely reject H_0 .

It should be noted that α_2 will be: $\text{Prob}(\text{reject } H_0 | H_0 \text{ is true, conditions (i) and (ii) are not rejected in step 1})$

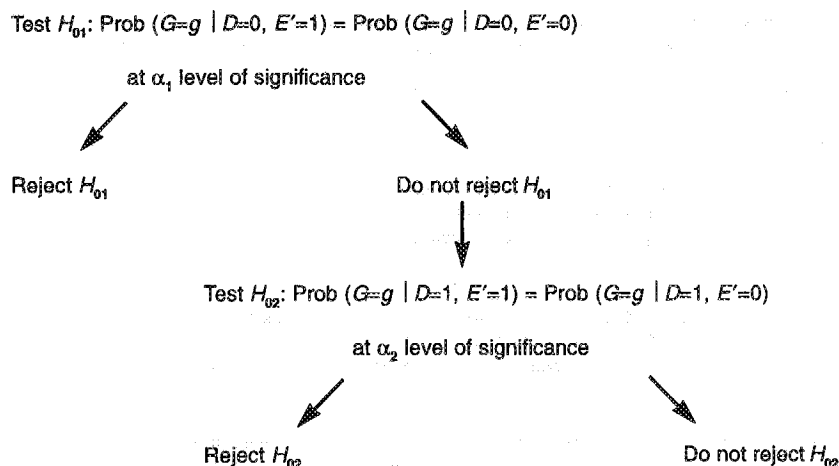


FIGURE 1. Two-step test for gene-environment interaction in the presence of misclassification of the environmental exposure. As discussed in the text, the two-step test should be used in addition to the conventional test for interaction.

only when conditions (i) and (ii) are true. The probability of rejecting H_0 without conditioning on failure to reject in step 1 will be $(1 - \alpha_1) \alpha_2$, i.e., the probability of not rejecting in step 1 \times the probability of rejecting in step 2, since both tests are independent.

We believe that a minimal requirement that any test of homogeneity designed to account for misclassification should possess is that the test should not falsely reject more often than its nominal alpha-level when there is no misclassification, even if a genotype-exposure association exists in the population (i.e., if condition (i) is false). Unfortunately, our two-step test with levels α_1 and α_2 does not satisfy this criteria since, under these conditions, it might reject more than α_2 percent of the time. That is, in situations where the environmental exposure is correctly measured, the genotype is associated with the environmental exposure among the controls and there is no true interaction, our two-step test may wrongly reject H_0 of homogeneity more than α_2 percent of the time, when the first step test has low power to detect a genotype-exposure association among the controls. As an example, table 1 shows the expected cell counts from a study where there is no misclassification, the environmental exposure is associated with the genotype among the controls and there is no interaction.

TABLE 1. Expected cell counts from a case-control study of a gene-environment interaction

Exposure	Cases			Controls		
	Genotype			Genotype		
	+	-	Combined	+	-	Combined
Yes	80	40	120	10	60	70
No	40	40	80	10	120	130
Combined	120	80	200	20	180	200

The power to detect an association in the control group is approximately 32 percent and the power to detect an association in the case group is approximately 65 percent. Thus, the probability of failing to reject in the first step but rejecting at the second step will be $0.68 \times 0.65 = 0.44$ which will lead us to frequent false positive conclusions about H_0 . On the other hand, since there is no misclassification, the type I error of a conventional test for interaction would be its nominal alpha-level.

Note that, when we have both misclassification and an exposure-disease association in the control population which we fail to detect in the first step due to low power, neither the conventional test nor the two-step test for interaction will have the stated alpha-level. To protect us from falsely rejecting H_0 of homogeneity a high percentage of the time both in this situation and in the situation illustrated in the example, we recommend the following *summary test procedure*: perform both a two-step test for interaction with alpha-levels α_1 and α_2 and a conventional α_2 -level test for interaction, and conclude rejection of H_0 only if *both* tests reject at α_2 -level. The possible outcomes and interpretations of the summary test procedure are shown in table 2. This summary test will be conservative in the following sense. In the absence of misclassification, the type I error of the conventional test is equal to its stated alpha-level (α_2). Similarly, when there is misclassification but conditions (i) and (ii) hold, then the type I error of the conventional and the two-step test is equal to their stated alpha-level (α_2). Thus, in either of these cases, the type I error of the summary test procedure will be no greater than α_2 . Finally, if misclassification is present and conditions (i) and (ii) do *not* hold, then we can place no upper bound on the type I error in the summary test procedure; however,

TABLE 2. Suggested decision matrix for the summary test procedure for testing the H_0 of no multiplicative gene-environment interaction when the presence of differential misclassification is suspected

Outcome of TST*,‡	Outcome of CT*,†	
	Reject H_0	Do not reject H_0
Reject H_0	Reject H_0 with less than α_2 type I error.§ The estimated interaction parameter is likely to be biased toward the null	Do not reject H_0 : Rejection by TST may be due to the presence of an exposure-genotype association not detected in the first step of TST
Do not reject H_0	Do not reject H_0 : Disagreement between CT and TST could be due to the effects of misclassification or low power of TST	Do not reject H_0 : An interaction is either not present or is of too small magnitude to be detected in the current study
Reject H_0 in first step of TST	Reject that conditions (i) and (ii) are true with α_1 type I error. We do not know whether rejection by CT is due to the effects of misclassification or to the presence of a true interaction	Reject that conditions (i) and (ii) are true with α_1 type I error. We do not know whether failure to reject by CT is due to the effects of misclassification, the absence of a true interaction, or lack of power

* CT, conventional test; TST, two-step test.

† Conventional test of H_0 of no multiplicative interaction with α_2 type I error.

‡ Two-step test of H_0 of no multiplicative interaction with α_1 and α_2 type I errors, for first and second steps, respectively.

§ Note that it is still possible that we made a type II error in the first step test of TST, in which case the type I error of the summary test procedure would not necessarily be less than α_2 .

this type I error will be smaller than if we would either conduct the conventional test or the two-step test alone.

EXAMPLE

We illustrate the implications of our findings with the analysis of data from a case-control study of light to moderate former and current smokers (lifetime smoking dose less than 40 pack-years). Our goal is to assess the association between smoking-induced lung cancer and a deletion of the *GSTM1* gene. The *GSTM1* gene is responsible for the glutathione-S-transferase M1 activity which is involved in the detoxification of tobacco carcinogens. When both copies of this gene are deleted (30–60 percent of the population depending on the ethnic background) (9), the detoxifying glutathione-S-transferase M1 activity is not expressed and, as a consequence, subjects with this deletion might have an increased susceptibility to develop smoking-induced lung cancer. The objective of this analysis is to assess whether the odds ratio for the *GSTM1* deletion differs according to the level of fruit and yellow vegetable intake, since antioxidant vitamins or other products found in these foods can reduce oxidative damage from tobacco carcinogens. The purpose of this analysis is to illustrate a methodological point, and no other inferences should be made from the analysis.

To evaluate the potential effect modification by the level of fruit intake, we used dietary information ob-

tained from a food frequency questionnaire which estimates the level of food intake during the year before diagnosis for cases and the year before enrollment for controls. This information is used as a surrogate for diet during the etiologically relevant period of exposure which may occur many years before diagnosis. This surrogate measure is subject to differential misclassification since the ability to recall past diet is likely to differ for cases and controls, and cases, but not controls, might have changed their diets due to the progression of their disease (10). For the sake of simplicity, stratification by ethnicity, smoking habits, or other potential confounders will be ignored in this example.

The observed data from this case-control study is presented in table 3. These data suggest that a complete deletion of the *GSTM1* gene increases the risk of lung cancer about six times among subjects with low intakes of fruits, and has no effect among subjects with higher intakes (Mantel-Haenszel test for interaction $\chi^2_{(1)} = 7.14$, $p = 0.01$). Similarly, a deletion in the *GSTM1* gene is associated with a fourfold increase in risk among subjects with low intakes of yellow vegetables but not among subjects with higher intakes (Mantel-Haenszel test for interaction $\chi^2_{(1)} = 4.56$, $p = 0.03$). Another interpretation of these results, however, would be that the observed heterogeneity of the stratum-specific odds ratios is merely the result of a large amount of misclassification in the assessment of fruit and vegetable intake.

TABLE 3. Examples of observed data from a case-control study of lung cancer and a defect in the *GSTM1* gene* stratified by the level of fruit and yellow vegetable intake, with odds ratios (ORs) and 95 percent confidence intervals (CIs)

Example 1: Interaction between fruit intake and <i>GSTM1</i> genotype				
	Low intake of fruits†		High intake of fruits†	
	<i>GSTM1</i> -	<i>GSTM1</i> +	<i>GSTM1</i> -	<i>GSTM1</i> +
Cases	16	4	32	25
Controls	33	52	98	80
	OR _{E=1} = 6.30		OR _{E=0} = 1.04	
	95% CI 2.01-19.49		95% CI 0.58-1.89	

Example 2: Interaction between yellow vegetable intake and <i>GSTM1</i> genotype				
	Low intake of yellow vegetables†		High intake of yellow vegetables†	
	<i>GSTM1</i> -	<i>GSTM1</i> +	<i>GSTM1</i> -	<i>GSTM1</i> +
Cases	19	5	29	24
Controls	38	44	93 ^a	88
	OR _{E=0} = 4.40		OR _{E=1} = 1.14	
	95% CI 1.54-12.44		95% CI 0.62-2.10	

* *GSTM1*+, deletion is present; *GSTM1*-, deletion is not present.

† The 33.3 percentile of the distribution of fruit intake among controls was used as the cut-off point.

‡ E', level of misclassified environmental exposure (see text).

According to the findings presented in the methods section, if 1) the true food intake is independent of the *GSTM1* genotype among the controls, and 2) the misclassification of food intake is non-differential with respect to the *GSTM1* genotype, then a test of the null hypothesis of homogeneity of the stratum-specific odds ratios will be valid even in the presence of differential misclassification of food intake. Therefore, if these conditions are satisfied in our data, we would be able to distinguish between the two possible interpretations of the results.

To test for homogeneity, we will use the two-step test presented in the previous section with $\alpha_1 = 0.05$ and $\alpha_2 = 0.05$.

Example 1: Test for interaction between fruit intake and *GSTM1* genotype

Step 1 H_{01} : *GSTM1* deletion is independent of fruit intake among the controls. The test statistic for this test is $\chi^2_{(1)} = 6.06$, $p = 0.01$, therefore H_{01} is rejected with a 5 percent type I error. The odds ratio for the *GSTM1*-fruit intake association among the controls is 1.93 (95 percent confidence interval (CI) 1.14-3.26).

Since H_{01} has been rejected, we will conclude that conditions for validity of a test for homogeneity are likely to be false and, therefore, we cannot report a valid p -value for the test for homogeneity.

Example 2: Test for interaction between vegetable intake and *GSTM1* genotype

Step 1 H_{01} : *GSTM1* deletion is independent of yellow vegetable intake among the controls. The test statistic for this test is $\chi^2_{(1)} = 0.57$, $p = 0.45$. Since H_{01} cannot be rejected at a 5 percent level of significance, we will proceed to the second step. The odds ratio for the *GSTM1*-yellow vegetable intake association among the controls is 1.22 (95 percent CI 0.73-2.06).

Step 2 H_{02} : *GSTM1* deletion is independent of yellow vegetable intake among the cases. The test statistic for this test is $\chi^2_{(1)} = 4.21$, $p = 0.04$ and the odds ratio is 0.32 (95 percent CI 0.11-0.95). Thus, the p -value for the two-step test for homogeneity is 0.04 and the null hypothesis of homogeneity of the stratum-specific odds ratios is rejected with a conditional 5 percent type I error.

Since, as reported above, the Mantel-Haenszel test for interaction also rejected H_0 at the 5 percent level of significance ($\chi^2_{(1)} = 4.56$, $p = 0.03$), we conclude rejection of H_0 with a less than 5 percent type I error. It should be noted, however, that the estimated stratum-specific odds ratios are still affected by differential misclassification. According to the second result of the paper, if conditions (i) and (ii) are true as was concluded in the first step test, the estimated interaction parameter or ratio of stratum-specific odds ratios (0.26, 95 percent CI 0.08-0.90) will be biased toward the null value.

SUMMARY AND CONCLUSIONS

In this paper we have identified a set of conditions under which differential misclassification of the environmental exposure with respect to disease 1) does not introduce bias in the ratio of the stratum-specific odds ratios when a true interaction is not present and 2) underestimates the interaction parameter when a true interaction is present. The conditions are that (i) the true environmental exposure is independent of the genotype among the controls, and (ii) misclassification of the environmental exposure is non-differential with respect to the genotype. We have argued that these conditions are, a priori, likely to be satisfied in case-control studies of gene-environment interactions. We believe that this is true especially for condition (ii) because individual genotypes are unlikely to substantially affect the quality of the information collected on environmental factors. On the other hand, genotypes are often unrelated to environmental exposures conditional on potential confounders such as ethnicity.

Piegorsch et al. (8) showed that if we assume that the environmental exposure and the genotype are independent (our condition (i)), then a test of the independence of the environmental exposure and the genotype among the cases can be used as a test for multiplicative interaction, and this test has better

power than a conventional test for interaction. In the presence of misclassification of the environmental exposure, we would also need to assume that condition (ii) is satisfied in order to have a valid test for interaction. However, in many instances we cannot be sure that conditions (i) and (ii) are true and, thus, it is important to be able to test the validity of these conditions. We have shown that these conditions can be tested among the controls using a χ^2 test of the null hypothesis of independence of the misclassified environmental exposure and the genotype. Therefore, we proposed a test for homogeneity which consists of two independent tests: First, we test for the genotype-exposure association among the *controls* and, if we cannot reject this null hypothesis, we proceed to the second step and conduct a test for the genotype-exposure association among the *cases*. This latter test is interpreted as a test of homogeneity of the true odds ratios. Our two-step test for multiplicative interaction has the stated alpha-level conditional on proceeding to the second step *only* when conditions (i) and (ii) are true. Thus, when there is a genotype-exposure association in the control population, but the power to detect this association is small, the two-step test might reject the null hypothesis of interaction a higher percentage of the time than the nominal alpha-level. Therefore, we recommend a *summary test procedure*, where one performs both the two-step test and a conventional test for interaction and then conclude rejection of the H_0 of homogeneity only if both tests reject.

The advantages of this summary test procedure are that we have better control over the type I error when we conclude rejection of H_0 of homogeneity, we can be more confident that the true odds ratios are homogeneous when we fail to reject H_0 in both tests, and we can distinguish between situations where differential misclassification is a more or less likely explanation for a rejection of H_0 by the conventional test (table 2). On the other hand, the first step test allows us to identify situations where the interaction parameter is likely to be biased toward the null value, in the sense that if we fail to reject in the first step and conditions (i) and (ii) are true, the interaction parameter estimated from the data is biased toward the null. Finally, the first step of the two-step test can help determine settings in which conditions (i) and (ii) are false. This is important since one reason why they could be false is due to an unexpected gene-environment association among the controls, which may reflect the presence of confounding by possible poorly measured factors such as ethnicity, problems in the selection of controls, or errors in the environmental exposure that are differential by the genotype.

The price we are paying for having a more informative test of the hypothesis H_0 of homogeneity is that α_1 percent of the times (where α_1 is the type I error in the first step test) we will not proceed to the second step and make no conclusion about H_0 although conditions (i) and (ii) are satisfied. If we have particularly strong a priori reasons to believe that conditions (i) and (ii) are true in a given study, we should increase α_1 in order to reduce the probability of making no conclusions about H_0 when conditions (i) and (ii) are in fact satisfied. Another cost of the summary test procedure relative to the conventional test is that by making it more difficult to reject H_0 , we are not only reducing the type I error but also increasing the type II error, and, therefore, the power of the test will be reduced. We believe that this loss of statistical power is offset by the benefits derived from our method. The finding of an interaction often has important scientific implications and, thus, reducing the probability of making false conclusions about the presence of interactions may be justified. When our test fails to reject the null hypothesis of no interaction, we do not conclude that there is no interaction, but we rather see it as a warning that we should get better data with less misclassification or learn about the misclassification matrix if we want to make a conclusion about the presence of an interaction, since the data we have does not provide enough evidence to make a conclusion.

It should be noted that, in situations where we have information about the misclassification matrix, methods to correct for misclassification of exposure might provide us with a more powerful test for interaction than the one recommended in this paper. Moreover, whereas our results indicate the direction in which the stratum-specific odds ratios will be biased when there is true interaction and conditions (i) and (ii) are true, additional information on misclassification probabilities might enable us to compute corrected estimates of effect (3, 4). Finally, it is important to realize that the methods presented in this paper cannot be extended to provide tests of additive genotype-environment interactions in the presence of differential misclassification.

In summary, our findings provide the researcher with some guidance as how to deal with differential misclassification in the context of case-control studies of multiplicative genotype-environment interactions when information on the misclassification matrix is not available.

ACKNOWLEDGMENTS

Supported by National Institutes of Health grant no. ES/CA PO1 06409. Dr. García-Closas was supported by a

scholarship from the Real Colegio Complutense, Cambridge, MA.

The authors thank Dr. Jeromin Zettelmeyer for his helpful comments on the paper.

REFERENCES

1. MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston, MA: Little Brown, 1970.
2. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112:564-9.
3. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York, NY: Van Nostrand Reinhold, 1982.
4. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. *Am J Epidemiol* 1989;129:1062-71.
5. Copeland KT, Checkoway H, McMichael AJ, et al. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488-95.
6. Birkett NJ. Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. *Am J Epidemiol* 1992;136:356-62.
7. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71-6.
8. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med* 1994;13:153-62.
9. Lin HJ, Han CY, Bernstein DA, et al. Ethnic distribution of the glutathione transferase Mu 1-1 (*GSTM1*) null genotype in 1473 individuals and application to bladder cancer susceptibility. *Carcinogenesis* 1994;15:1077-81.
10. Willett W. Recall of remote diet. In: Nutritional epidemiology. Monographs in epidemiology and biostatistics, vol 15. New York, NY: Oxford University Press, 1990:127-32.